Soumya Mukherjee | CH24M571 |

Lab Report Question 1

1.0 Preface

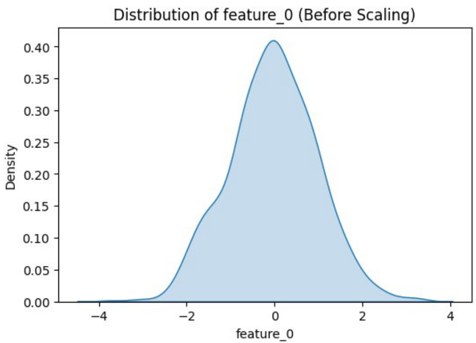
The question 1 of the assignment asks us to do the following :

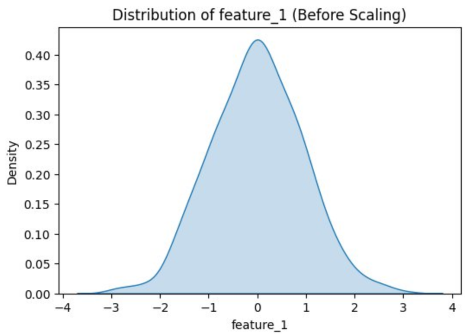
- Ridge Regression with default hyper parameters
- Lasso Regression with default hyper parameters
- Discussion of the impact of regularization term alpha on the coefficient's values and the model's performance
- Polynomial Regression of increasing degree 3-10 to calculate the MSEs and discuss the bias , variance , over fitting and the under fitting
- Kernel Ridge Regression with alpha = 1

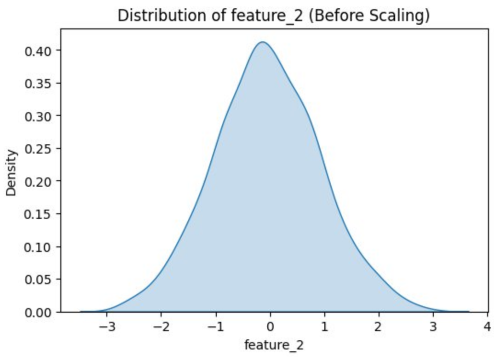
2.0 Methodology :

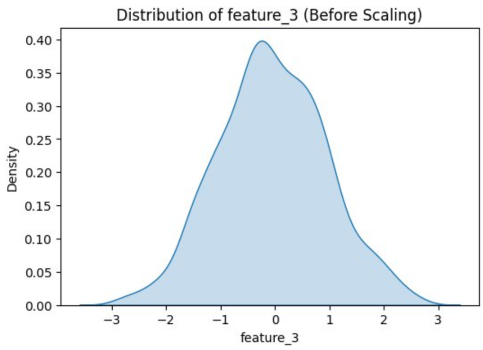
2.0.1 Data Pre-processing :

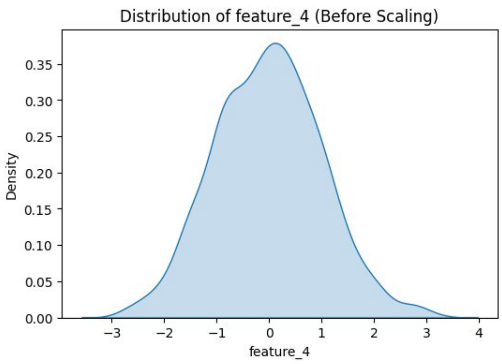
- The dataset was checked for null values to ensure data quality . Results show no null or missing values .
- Scatter plots and Kernel Density Estimation (KDE) plots were used to understand the distribution and relationships in the data. The features had similar scales .

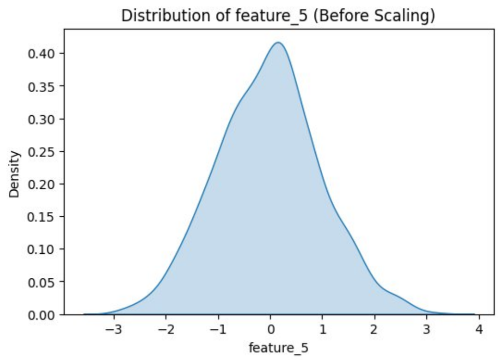


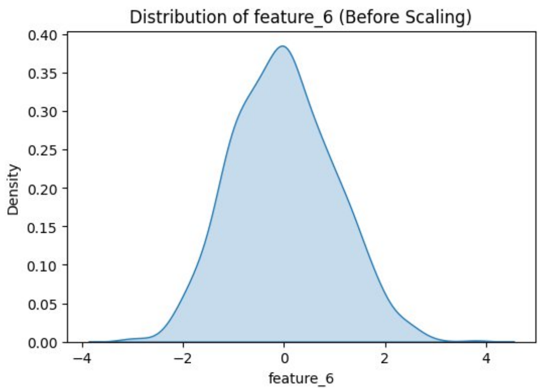


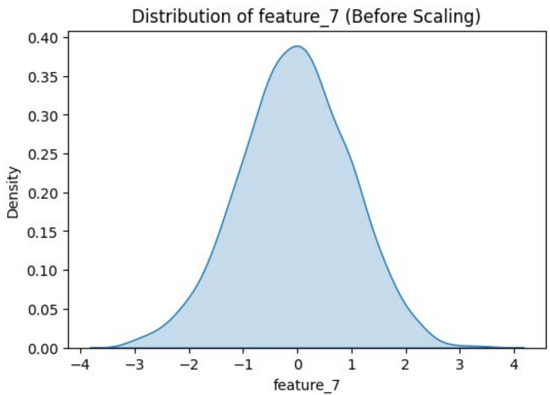

Distribution of feature_2 (Before Scaling)
Density
feature_2

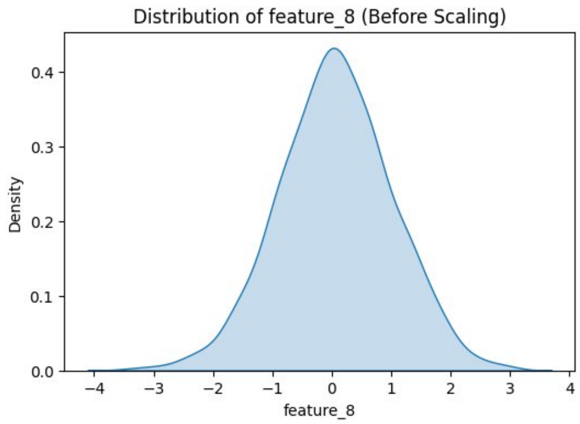

Distribution of feature_3 (Before Scaling)
Density
feature_3

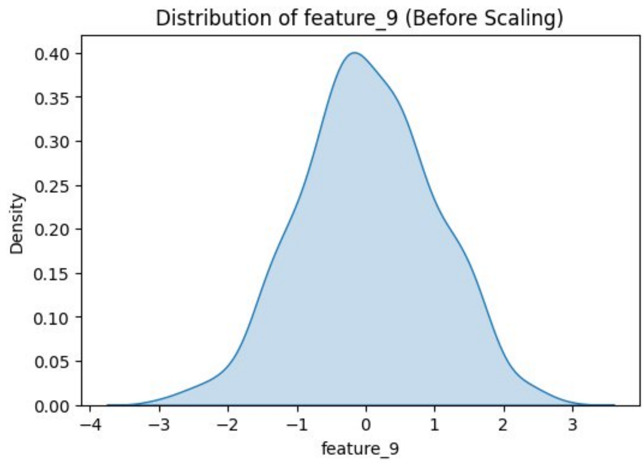

Distribution of feature_4 (Before Scaling)
Density
feature_4

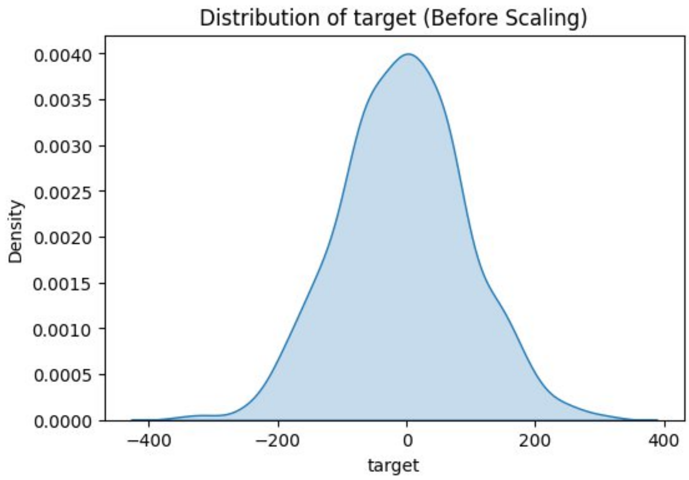

Distribution of feature_5 (Before Scaling)
Density
feature_5


Distribution of feature_6 (Before Scaling)
Density
feature_6


Distribution of feature_7 (Before Scaling)
Density
feature_7







The goal of the KDE plot was to understand the distribution of the data . Scatter plot was performed on all of the features to understand the relationships in the data

- Data was split into 80:20 ratio by using train-test split
- Feature Scaling was done using Standard Scalar to ensure all features contribute equally to the model . This is important for the Kernel based methods

Effects of Feature scaling on the methods asked in the question :

– Ridge and Lasso Regression : Regularization terms are sensitive to feature magnitudes. Scaling ensures that the regularization penalty is applied uniformly.

- Polynomial Regression: Scaling prevents numerical instability when creating higher-degree polynomial features.
- Kernel Ridge Regression: Kernels like RBF and polynomial are sensitive to feature scales. Scaling ensures that distances and similarities are calculated correctly.

2.0.2 Feature Engineering :

Polynomial features of degree 3 were created using PolynomialFeatures from ‘sklearn’ library. This allows the model to capture non-linear relationships in the data .

2.0.3 Model Implementation details :

- Ridge and Lasso Regression: Implemented with default hyperparameters and polynomial features of degree 3
- Polynomial Regression: Models of degrees 3–10 were created, and their mean squared errors (MSEs) were calculated.
- Kernel Ridge Regression: Implemented with alpha = 1 and evaluated using different kernels (linear, poly, and rbf).

3.0 Results and Analysis

3.1 Ridge Regression :

The Ridge regression model with polynomial features of degree 3 achieved an MSE of 0.757 on the test data. Ridge regression applies L2 regularization, which shrinks coefficients but does not eliminate them entirely. The moderate MSE indicates that the model generalizes reasonably well but may still have some bias due to regularization.

3.2 Lasso Regression

The Lasso regression model with polynomial features of degree 3 achieved an MSE of 0.120 on the test data. Lasso regression applies L1 regularization, which can shrink some coefficients to zero, effectively performing feature selection.The lower MSE compared to Ridge regression suggests that Lasso is better at identifying and ignoring irrelevant features, leading to improved generalization.

3.3 Comparison of Ridge and Lasso Regression :

Impact of Regularization:

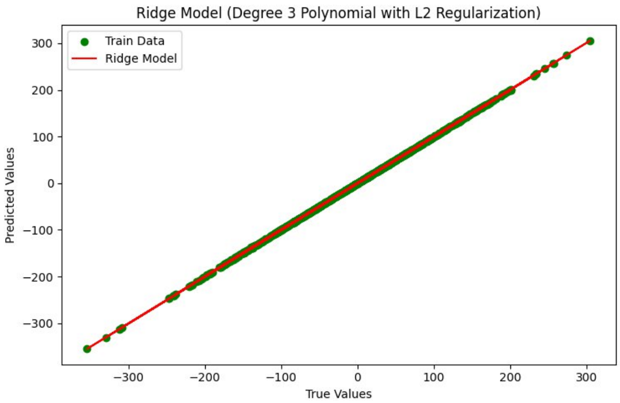
Ridge Regression: Shrinks coefficients proportionally, retaining all features but reducing their impact.

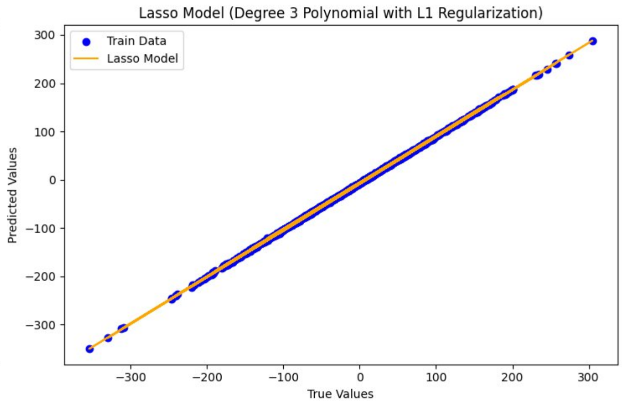
Lasso Regression: Shrinks some coefficients to zero, effectively performing feature selection.

Performance:

Lasso regression outperformed Ridge regression, achieving a significantly lower MSE (0.120 vs. 0.757).

This suggests that the dataset contains irrelevant or less important features that Lasso effectively ignored.

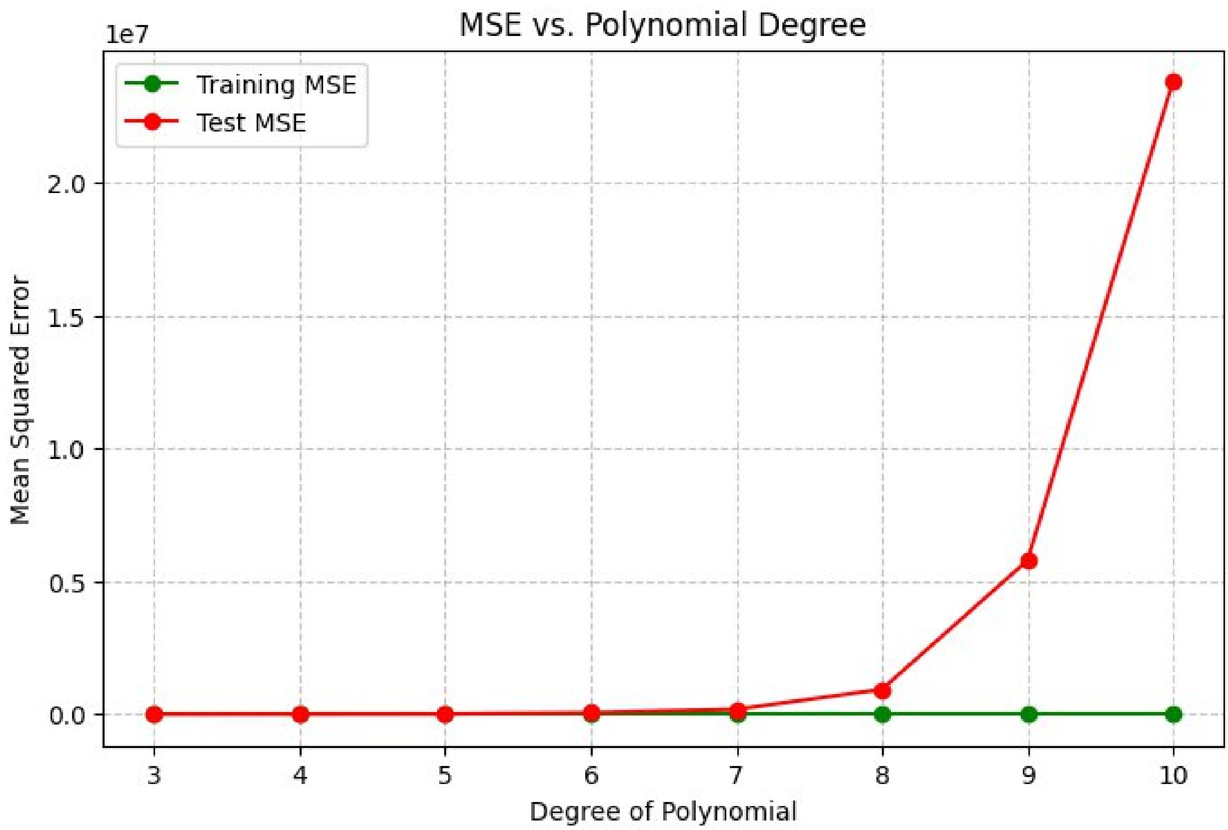




3.4 Polynomial Regression (Degrees 3–10)

Degree	Train MSE	Test MSE
3	0.006	0.020
4	3.276e-25	2755.438
5	3.612e-25	7124.701
6	1.676e-24	50671.999
7	4.170e-24	176819.395
8	4.442e-23	929230.077
9	2.398e-22	5778600.442
10	3.088e-21	23795494.855

The MSE values for polynomial regression models of degrees 3–10 are as above :



Analysis:

As the degree of the polynomial increases, the train MSE decreases significantly, approaching zero. This indicates that the model is fitting the training data almost perfectly.

However, the test MSE increases dramatically, indicating severe over-fitting. Higher-degree polynomials capture noise in the training data, leading to poor generalization.

The bias-variance trade-off is evident:

Low-degree models (e.g., degree 3): Higher bias but lower variance. They generalize better but may underfit the data.High-degree models (e.g., degree 10): Lower bias but extremely high variance. They overfit the training data and fail to generalize.

For this dataset, a polynomial regression model with a degree of 3–4 is likely the best choice, as it balances bias and variance and generalizes well to unseen data.Beyond degree 4, the test MSE increases sharply, indicating that higher-degree polynomials are overfitting the training data.

3.5 Kernel Ridge Regression

The MSE values for kernel ridge regression with different kernels are as follows:

Kernel	Train MSE	Test MSE
Linear	50.2971	50.1510
Poly	0.9057	1.9265

RBF	326.3587	799.6944

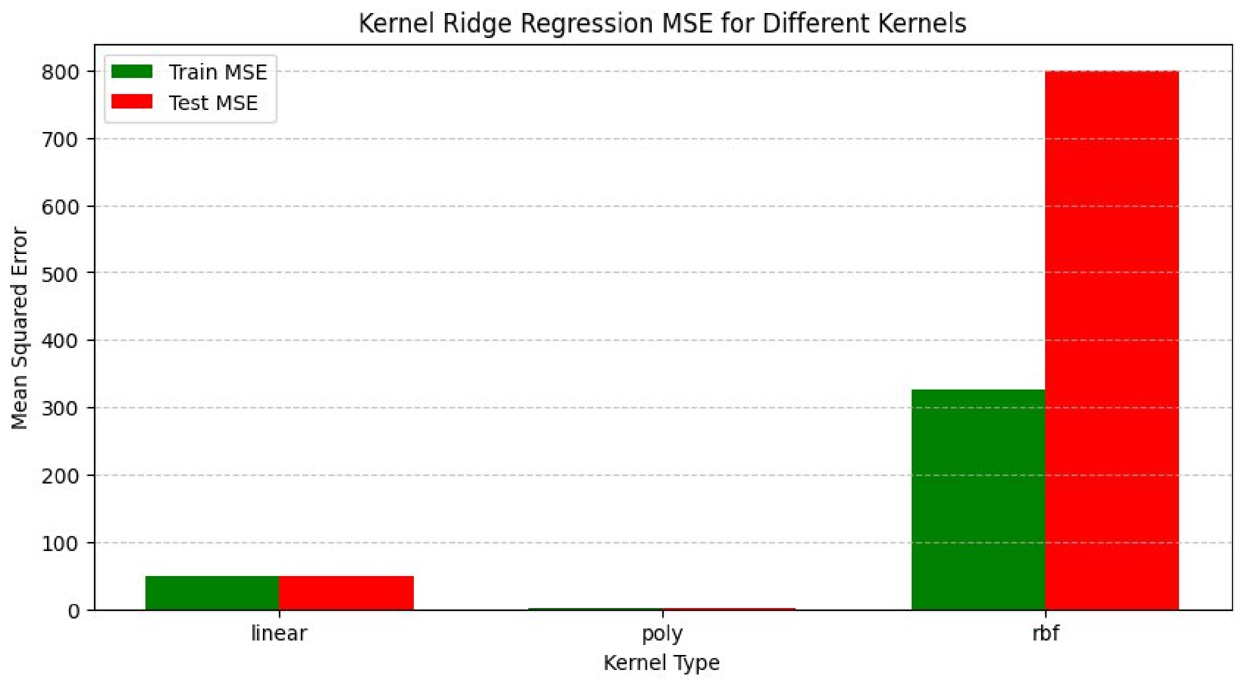
Analysis:

Linear Kernel: High train and test MSE values indicate underfitting. The linear kernel is too simple to capture the underlying patterns in the data.

Polynomial Kernel: Low train MSE but higher test MSE suggests overfitting. The polynomial kernel captures the training data well but fails to generalize to unseen data.

RBF Kernel: High train and test MSE values indicate poor performance. The RBF kernel is too complex for this dataset, leading to overfitting and poor generalization.

Conclusion: None of the kernels performed exceptionally well, but the polynomial kernel showed the best balance between train and test MSE.The RBF kernel should be avoided unless we allow for further tuning (e.g., adjusting the gamma parameter) is performed to control its complexity. The linear kernel is too simple for this dataset and should not be used unless the data is inherently linear.



4.0 Conclusion

Ridge vs. Lasso: Lasso regression outperformed Ridge regression, likely due to its ability to perform feature selection.

Polynomial Regression: Higher-degree polynomials led to severe overfitting, highlighting the importance of balancing bias and variance.

Kernel Ridge Regression: The polynomial kernel showed the best performance among the tested kernels, but further tuning of hyperparameters (e.g., alpha, degree) could improve results.

Scaling: Data scaling played a critical role in ensuring fair and accurate model performance across all methods.

Lab Report Question 2

1.0 Preface

The question 2 of the assignment asks us to do the following :

- Classify the dataset using Logistic regression and find summary statistics
- Implement the LDA and QDA and evaluate the model performance
- Implement stochastic gradient descent classifier from scratch

2.0 Methodology :

2.0.1 Data Pre-processing :

- We check the data for null column count and use df.describe to get an idea about the data
- Since no columns with categorical variables are present , we can can proceed without encoding the data
- The dataset was checked for missing values, and no missing entries were found, so no imputation was required.
- Since Logistic Regression, LDA, QDA, and Stochastic Gradient Descent (SGD) are sensitive to feature scales, the data was standardized to have zero mean and unit variance using StandardScaler to ensure that all features contribute equally to the model.
- The dataset was divided into training (80%) and testing (20%) sets to evaluate model performance on unseen data.

2.0.2 Feature Engineering :

Feature Importance Analysis:

Logistic Regression provides feature coefficients that indicate the importance of each feature.

The most important features identified were:

feature_0 , feature_3 ,feature_7

Summary Statistics for Important Features:

Mean, standard deviation, minimum, and maximum values for these features were computed to understand their distribution. For instance:

feature_0 had a mean of 0.177 and a standard deviation of 1.530.

feature_3 had a mean of 0.033 and a standard deviation of 1.591.

feature_7 had a mean of -0.001 and a standard deviation of 1.465.

3.0 Results and Analysis

3.0.1 Logistic Regression Performance

Accuracy: 93%

Precision, Recall, and F1-Score: Achieved > 0.92 across both classes (0 and 1).

The high and balanced precision and recall indicate that the model effectively differentiates between the two classes without significant bias.

3.2 LDA and QDA Performance:

Accuracy for Both Models: 94.5%

The consistency in performance across Logistic Regression, LDA, and QDA suggests that the data may be linearly separable or near-linearly separable.

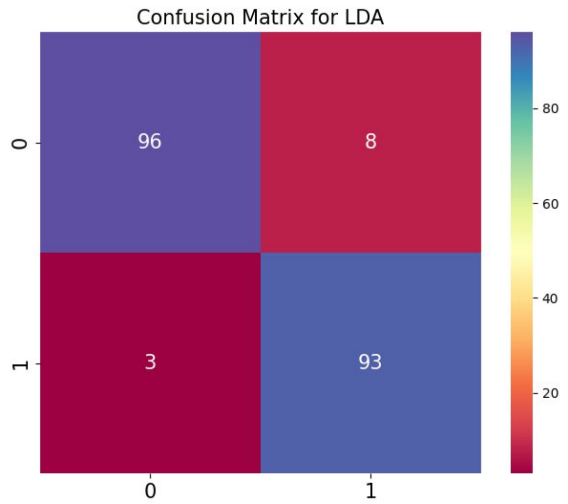
3.3 Stochastic Gradient Descent (SGD) Classifier :

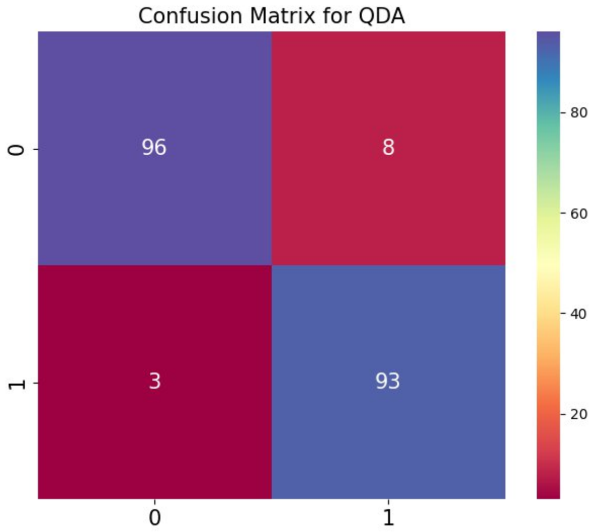
Best Learning Rate: 0.001 (based on log-loss minimization).

Accuracy: 93% (same as other models).

Analysis: The effectiveness of the SGD classifier with a custom implementation indicates that the dataset's features are well-suited for linear decision boundaries. The log-loss was minimized effectively with a low learning rate, ensuring stable updates during training.

Confusion Matrix for LDA and QDA :





Conclusion:

- Consistent Performance: All models, including the custom SGD implementation, achieved 94.5% accuracy, indicating that the dataset is linearly separable and well-suited for these classifiers.
- Importance of Feature Scaling: Standardization contributed to stable and efficient training, especially for SGD, which is sensitive to feature scales.
- Feature Importance: *feature_0, feature_3, and feature_7* were identified as the most significant features, suggesting they contain most of the predictive information.